

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:09:29 ; Search time 299.73 seconds
(without alignments)
23.463 Million cell updates/sec

Title: US-09-331-631a-7_COPY_81_140
Perfect score: 342
Sequence: 1 LORQYQCCGRCQEQDQGOR.....HENYHNHKKNRSEEGQOR 60

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	342	100.0	525	10 Q43358	Q43358 theobroma c
2	143.5	42.0	593	10 Q9SEW4	Q9SEW4 juglans reg
3	130.5	38.2	666	10 Q9SP14	Q9SP14 macadamia i
4	127.5	37.3	625	10 Q9SP13	Q9SP13 macadamia i
5	126.5	37.0	666	10 Q9SP15	Q9SP15 macadamia i
6	113	33.0	2123	5 Q909S7	Q909S7 dictyostell
7	104	30.4	810	10 Q9ZW13	Q9ZW13 cucurbita m
8	103.5	30.3	1038	5 Q60983	Q60983 dictyostell
9	101.5	29.7	304	4 Q15409	Q15409 homo sapien
10	100.5	29.4	1556	5 Q9U6A1	Q9U6A1 drosophila
11	100	29.2	523	5 Q9U1K4	Q9U1K4 drosophila
12	100	29.2	666	5 Q9W318	Q9W318 drosophila
13	96	28.1	171	11 Q61118	Q61118 mus musculu
14	94.5	27.6	420	11 Q35861	Q35861 mus musculu
15	93.5	27.3	828	5 Q9VZU6	Q9VZU6 drosophila
16	93	27.2	930	10 Q9SZY9	Q9SZY9 arabidopsis
17	92.5	27.0	838	5 Q9YOC9	Q9YOC9 dictyostell
18	92	26.9	139	11 Q62006	Q62006 mus musculu
19	92	26.9	454	5 P91642	P91642 drosophila

20	92	26.9	531	5 Q9VT57	Q9VT57 drosophila
21	92	26.9	2074	11 Q88542	Q88542 mus musculu
22	91.5	26.8	517	5 Q9W4E8	Q9W4E8 drosophila
23	91	26.6	1457	5 Q44011	Q44011 dictyostell
24	91	26.6	3502	5 Q9VJ39	Q9VJ39 drosophila
25	90.5	26.5	776	5 Q9VFR7	Q9VFR7 drosophila
26	90	26.3	1094	5 Q23915	Q23915 dictyostell
27	89	26.0	259	5 Q9W483	Q9W483 drosophila
28	89	26.0	406	2 Q87306	Q87306 borrelia bu
29	89	26.0	646	2 Q9KW25	Q9KW25 vibrio chol
30	89	26.0	821	5 Q9W3A1	Q9W3A1 drosophila
31	89	26.0	833	4 Q9NTG8	Q9NTG8 homo sapien
32	89	26.0	1010	5 Q18346	Q18346 drosophila
33	89	26.0	2678	5 Q9NDS4	Q9NDS4 dictyostell
34	88.5	25.9	1326	5 Q9W4N4	Q9W4N4 drosophila
35	88	25.7	242	10 Q81251	Q81251 zea mays su
36	88	25.7	242	10 Q9SBE6	Q9SBE6 zea mays su
37	88	25.7	697	5 Q9VZX2	Q9VZX2 drosophila
38	88	25.7	791	5 Q00841	Q00841 dictyostell
39	88	25.7	1502	4 Q9ULJ8	Q9ULJ8 homo sapien
40	88	25.7	2150	5 Q23863	Q23863 dictyostell
41	87	25.4	420	5 Q76853	Q76853 dictyostell
42	87	25.4	551	10 Q43607	Q43607 pinus dulc
43	86.5	25.3	57	4 Q9Y4J1	Q9Y4J1 homo sapien
44	86.5	25.3	341	5 Q9VR72	Q9VR72 drosophila
45	86.5	25.3	438	5 Q9W2K6	Q9W2K6 drosophila

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	525 AA.
Q43358	Q43358			
AC	Q43358			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DE	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	VICILIN PRECURSOR.			
GN	CSV.			
OS	Theobroma cacao (Cacao).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;			
OC	Malvales; Malvaceae; Theobroma.			
OX	NCBI_TaxID=3641;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-LEAVES:			
RX	MEDLINE=92288309; PubMed=1600151;			
RA	McHenry L., Frittz P.J.;			
RT	*Comparison of the structure and nucleotide sequences of vicilin genes			
RL	of cocoa and cotton raises questions about vicilin evolution.*;			
DR	Plant Mol. Biol. 18:1173-1176(1992).			
DR	EMBL; X62625; CAA44493.1; -.			
DR	EMBL; X62626; CAA44494.1; -.			
DR	HSSP; P02853; 2PHL.			
DR	MENDEL; 30919; Thecc; 1188; 30919.			
DR	INTERPRO: IPR001113; -.			
DR	PFAM; PF00546; Seedstore-7s; 1.			
DR	PRODOM; PD081059; -; 1.			
FT	SIGNAL.			
FT	POTENTIAL.			
FT	CHAIN 25 525 VICILIN.			
SO	SEQUENCE 525 AA; 60798 MW; 19114CD5C248905D CRC64;			

Query Match 100.0%; Score 342; DB 10; Length 525;
Best Local Similarity 100.0%; Pred. No. 2.2e-30;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LORQYQCCGRCQEQDQGORCKCWEQKCEDEGEHENYHNHKKNRSEEGQOR 60
DB 81 LORQYQCCGRCQEQDQGORCKCWEQKCEDEGEHENYHNHKKNRSEEGQOR 140

```

RESULT 2
ID Q9SEM4 PRELIMINARY: PRT: 593 AA.
AC Q9SEM4:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE VICILIN-LIKE PROTEIN PRECURSOR (FRAGMENT).
OS Juglans regia (English walnut).
OC Eukaryota; Viridiplantae; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fagales; Juglandaceae; Juglans.
OX NCBI_TaxID=51240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SUNLAND: TISSUE-SOMATIC EMBRYO LINE;
RA Teuber S.S., Jarvis K.C., Peterson W.R., Dandekar A.M., Ansari A.A.;
RT "Identification and cloning of a cDNA encoding a vicilin-like protein,
RT Jug r 2, from English walnut kernel (Juglans regia): a major food
RT allergen."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF066055; AAF18269.1; -.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
FT NON_TER 1
SQ SEQUENCE 593 AA: 69990 MW: 9BA127E19B18C0D8 CRC64;

```

```

Query Match 42.0%; Score 143.5; DB 10; Length 593;
Best Local Similarity 38.8%; Pred. No. 2.2e-08;
Matches 31; Conservative 10; Mismatches 16; Indels 23; Gaps 3;
QY 2 QRYOQCGRCRCEQGGQREOQCGRCCKWEQYKEQ-----ERG-----EH 41
DB 120 QOYHRCQRCQIQEOSPERRQOQRCERQYKEQGGREKPEASPRRESRGREEQQRH 179
DB 42 ENYHNK---KNKSEEEQO 58
DB 180 NPYFFHSQSRHSRESEGE 199

```

```

RESULT 3
ID Q9SPL4 PRELIMINARY: PRT: 666 AA.
AC Q9SPL4:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE VICILIN PRECURSOR.
GN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OX NCBI_TaxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUT KERNEL.
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;
RT "A family of antimicrobial peptides is produced by processing of a 7S
RT globulin protein in Macadamia integrifolia."
RL Plant J. 0:0-0(1999).
DR EMBL: AF161884; AAD54245.1; -.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
SQ SEQUENCE 666 AA: 78243 MW: 0ECA22F8710F8A7B CRC64;

```

```

Query Match 38.2%; Score 130.5; DB 10; Length 666;
Best Local Similarity 34.3%; Pred. No. 6.6e-07;

```

```

Matches 24; Conservative 12; Mismatches 23; Indels 11; Gaps 1;
QY 2 QRYOQCGRCRCEQGGQREOQCGRCCKWEQYKEQ-----WEQYKQDREGEHNYHNKKN 50
DB 123 QOYHRCQRCQIQEOSPERRQOQRCERQYKEQGGREKPEASPRRESRGREEQQRH 182
QY 51 RSEEEGGQR 60
DB 183 EDNKRDPQQR 192

```

```

RESULT 4
ID Q9SPL3 PRELIMINARY: PRT: 625 AA.
AC Q9SPL3:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE VICILIN PRECURSOR (FRAGMENT).
GN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OX NCBI_TaxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUT KERNEL.
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;
RT "A family of antimicrobial peptides is produced by processing of a 7S
RT globulin protein in Macadamia integrifolia."
RL Plant J. 0:0-0(1999).
DR EMBL: AF161885; AAD54246.1; -.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
FT NON_TER 1
SQ SEQUENCE 625 AA: 73586 MW: 415808A89D370296 CRC64;

```

```

Query Match 37.3%; Score 127.5; DB 10; Length 625;
Best Local Similarity 34.3%; Pred. No. 1.3e-06;
Matches 24; Conservative 13; Mismatches 22; Indels 11; Gaps 1;
QY 2 QRYOQCGRCRCEQGGQREOQCGRCCKWEQYKEQ-----WEQYKQDREGEHNYHNKKN 50
DB 82 QOYHRCQRCQIQEOSPERRQOQRCERQYKEQGGREKPEASPRRESRGREEQQRH 141
QY 51 RSEEEGGQR 60
DB 142 GDNKRDPQQR 151

```

```

RESULT 5
ID Q9SPL5 PRELIMINARY: PRT: 666 AA.
AC Q9SPL5:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE VICILIN PRECURSOR.
GN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OX NCBI_TaxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUT KERNEL.
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;
RT "A family of antimicrobial peptides is produced by processing of a 7S
RT globulin protein in Macadamia integrifolia kernels."
RL Plant J. 0:0-0(1999).
DR EMBL: AF161883; AAD54244.1; -.

```



```

DT 01-JAN-1998 (Tremblarel, 05, Last sequence update)
DT 01-NOV-1998 (Tremblarel, 08, Last annotation update)
DE CAGH44 (FRAGMENT).
GN CAGH44.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97369492; PubMed=9225980;
RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA Breschei T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT "CDNAs with long CAG trinucleotide repeats from human brain.";
RL Hum. Genet. 100:114-122(1997).
RN (2)
RP SEQUENCE OF 1-86 FROM N.A.
RA Mixx, P., Hinds, K., Suterer, C., Becker, M., Ozersky P.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE OF 1-86 FROM N.A.
RA Waterston R.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; U80741; AAB91439.1; -.
DR EMBL; AC003992; AAB96326.1; -.
FT NON_TER 304
FT SEQUENCE 304 AA; 34400 MW; DFB9D9E6BB877A1E CRC64;

```

	Query Match	29.7%	Score 101.5:	DB 4:	Length 304;
	Best Local Similarity	35.0%:	Pred No.	0.00054:	
	Matches	21; Conservative	20; Mismatches	16; Indels	3; Gaps
Oy	1 LORATOCQCGRCQEBOGGGREGDOCCCRKCMEDYKEDEGEHENVHNHKNRSEEEGQGR	60			
	: : : : : : : : : : : : :				
Db	LSI LAQQQQQQQQQ---QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQAHGCAAKAEQQQDQQQQQ	207			

RESULT	10
ID	0906A1
AC	0906A1
PRELIMINARY;	PRT; 1556 AA.
DT	01-MAY-2000 (TrEMBLrel_13, Created)
DT	01-MAY-2000 (TrEMBLrel_13, Last sequence update)
DT	01-MAY-2000 (TrEMBLrel_13, Last annotation update)
DE	HOMEODOMAIN TRANSCRIPTION FACTOR PROSPERO.
GN	PROS.
OS	Drosophila virilis (Fruit fly).
OC	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
OC	Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
OC	Ephyraoidea: Drosophilidae: Drosophila.
OX	NCBL_taxid=7244;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Xu C., Kaufmann R.C., Zhang J., Kladny S., Carthew R.W.:
RT	"Overlapping activators and repressors delimit transcriptional
RT	response to receptor tyrosine kinase signals in the Drosophila eye".
RL	Submitted (Sep-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF190405; AAF06660.1; --
KW	Homeobox: DNA-binding; Nuclear protein.
SQ	SEQUENCE 1556 AA: 17102 MW; 66FACEA2D73E644 CRC64;

Query Match	29.4%	Score 100.5	DB 5	Length 1556
Best Local Similarity	35.5%	Pred No. 0.0029		
Matches 22	Conservative 19	Mismatches 18	Indels 3	Gaps 1

```
OY      090Y0CCCGFCEEDOOOGREDOOCORCKMEWYKCOER-----GEHNHYNNHKKNRSEEEGO   58  
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
773    QQQQOQQQQQQQQQQQQQQQQQQEQQRRYEDEQEQRMMEOOOQLTRROOHLLQDLQOOO  83Z
```

Db 833 1: 834

RESULT 11

AC OQ01K4; 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
DE C2H2 ZINC FINGER TRANSCRIPTION FACTOR.
GN D-Spl.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Empidoidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON S;
RA Schrock F., Purnell B.A., Wimmer E.A., Jackie H.;
RT "Common and diverged functions of the Drosophila gene pair D-Spl and
RT buttonhead.";
RL Mech. Dev. 89:125-132(1999).
DR EMBL; AJ131022; CAB55429.1; -.
DR HSSP; P08047; ISP2.
DR INTERPRO; IPR000822; -.
DR Pfam; PF00096; zf-C2H2_3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_3.
SO SEQUENCE 523 AA; 54577 MW; 64FB8392A72A7A8B CRC64;

[illegible]

RESULT	12			
09W318				
ID	09W318	PRELIMINARY;	PRT;	666 AA.
DC	09W318			
DT	01-MAY-2000	(TrEMBLrel. 13, Created)		
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)		
DT	01-OCT-2000	(TrEMBLrel. 15, Last annotation update)		
DE	SPI PROTEIN.			
GN	Spi.			
OS	Drosophila melanogaster (fruit fly)			
CC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
CC	Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_Taxid=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY;			
RX	MEDLINE=20196006; PUBMED=10731132;			
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazel R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Ballif J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,			
RA	Balder R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Besson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			

